Proposed Ph.D. Synopsis

On

Early Detection and Classification of Respiratory Diseases Using Deep Learning

Submitted to

MGM University, Aurangabad(MS)-India.

For the Completion of Ph.D. Degree

In

Computer Science and Information Technology

Submitted by

Ms.Urvashi B. Deshmukh
Research Scholar
MGM University, Chhatrapati Sambhaji Nagar, India

UndertheGuidance

Dr.Prapti D .Deshmukh
Principal
Dr. G. Y. Pathrikar College of CS and IT.
MGM University, Chhatrapati Sambhaji Nagar

Introduction

Respiratory system: Respiratory system is structure of human body which is combination of lungs, airway, throat, mouth, diaphragm and throat and its purpose is to breath in and breath out. Statistical analysis indicates that lung disorders are associated with the majority of four out of top 10 causes of death which is shown through the chart with 10 deadliest diseases in the world with the comparison of deaths in year 2000 and 2015 as per data collected from [13].

The present-day lifestyle and diet are responsible for the increasing prevalence of lung and respiratory diseases. Lower respiratory infections, ranking as the fourth leading cause of mortality, stand as the deadliest communicable diseases globally [13].

Lung diseases represent one of the most prevalent medical situations globally, affecting millions. These conditions encompass a range of issues that impede the lungs' ability to function correctly. Medical imaging plays a pivotal role in diagnosing and managing diseases, offering non-invasive methods to visualize and analyse the internal structures of human body. Imaging techniques like Computed tomography (CT), magnetic resonance imaging (MRI), and X-rays are instrumental in detecting abnormalities monitoring disease progression, and guiding treatment strategies. Image data preprocessing has a vital role in enhancing accuracy & reliability of subsequent analyses.

Feature extraction is at the heart of medical image analysis transforming high-dimensional image data into meaningful representations for segmentation, classification, and detection tasks. Robust feature extraction enables precise identification of disease pattern, ensuring accuracy in diagnostics while minimizing subjective biases. Techniques like wavelet packet decomposition (WPD), Gray level co-occurrence matrix (GLCM), and statistical descriptors have been widely employed to extract relevant feature from dataset. For instance, wavelet transform capture localized frequency component, making them suitable for analysing intricate structures in medical images [1]. Similarly, GLMC based feature are effective in quantifying texture properties aiding in the differentiation of pathological and normal tissues [2]. Lung diseases such as lung cancer, Covid-19 and pneumonia, exhibit diverse pattern across varying severity level. The importance of extracting complementary features - statistical, textural, shape and edge based- has been emphasized in studies that demonstrate their potential in improving diagnostic outcome [3,4].

Research Problem:

Purpose of this perspective study is to evaluate chest features of lung diseases and correlate imaging techniques severity, the overall role of imaging technique in management of identifying the pattern and severity. It will take 5-15 min to read single image (CT/MRI/X-Ray) for doctor to overcome this manual detection we have automated the detection of severity and classify them into mild, moderate and severe class.

Objective:

- To evaluate the spectrum of features of lung diseases on chest using digital image processing techniques.

- To access the role of imaging techniques in diagnosis of lung diseases and affecting treatment plan
- To correlate the patterns found on imaging techniques and its severity using Deep learning techniques.

Literature review:

Stefanus Tao Hwa Kieu, Abdullah Bade et. al. “A Survey of Deep Learning for Lung Disease Detection on Medical Images: State-of-the-Art, Taxonomy, Issues and Future Directions” in this proposed system 98 articles publish from 2016-2020 consider to explain how deep learning is effective on disease detection in medical images. CNN, DBN, RNN etc where used as classification algorithms [5].

A. D. Gunasinghe, A. C. Aponso and H. Thirimanna, "Early Prediction of Lung Diseases," lung cancer is most common disease in the world comparative other type of cancer maximum no of death was caused by lung cancer, article includes the structural technologies for early detection of cancer such as Machine learning, deep learning and other [6].

Tianyi li, wei wei et. al. “Computer Aided Diagnosis of COVID-19 CT scan based on spatiotemporal Information Fusion” gave the more classification of COVID-19 CT scans and non-COVID-19 CT scans and analyzed the development of the patients’ condition through the CT scans. Gives accuracy, sensitivity and F1 score which takes 30 sec to calculate [7].

Yuanyuan Peng et. al. “Pulmonary Lobe Segmentation in CT Images Based on Lung Anatomy Knowledge” developed a system using developed in MATLAB and C++. To remedy the problem, they introduce a new framework based on lung anatomy knowledge for lung lobe segmentation. As result they develop a new lung lobe segmentation scheme that described method has a good performance in pulmonary fissure detection and lung lobe segmentation [8].

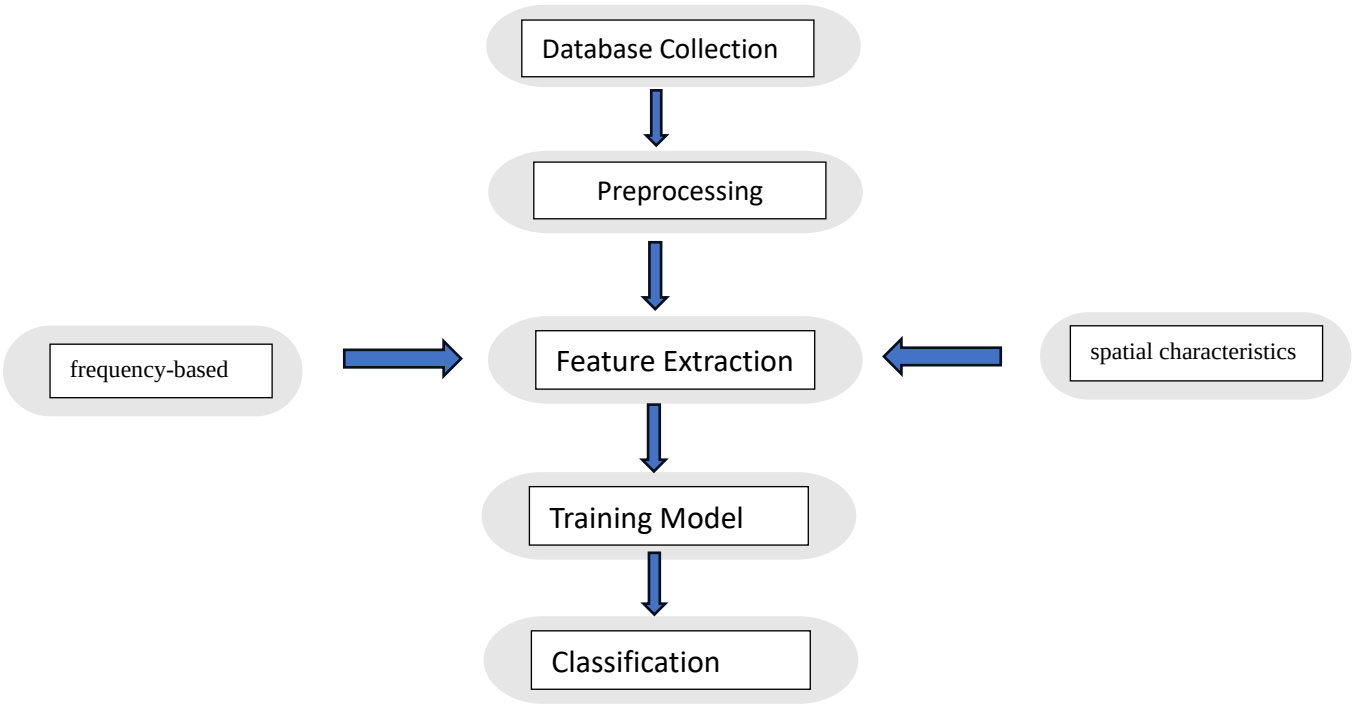
Suren Makajua, P.W.C. Prasad et al “Lung Cancer Detection using CT Scan Images” The proposed system is used to detect the cancerous nodule from the lung CT scan image using watershed segmentation for detection and SVM for classification of nodule as Malignant or benign using Image Processing Techniques [9].

X. Jin, Y. Zhang and Q. Jin, "Pulmonary Nodule Detection Based on CT Images Using Convolution Neural Network," In this system a image segmentation and region of interest extraction in pulmonary CT images the method suppresses the linear disturbance and removing a large number of false positive components when ensuring the accuracy of pulmonary nodule extraction [10].

Bakheet & Al-Hamadi, “Automatic Detection of COVID-19 Using Pruned GLCM-Based Texture Features and LDCRF Classification”, This study Presents a CAD system for COVID-19 detection using GLCM-based texture features and LDCRF classification, achieving superior diagnostic accuracy 95.88% [11].

Karthika M.S., Harikumar Rajaguru et al. “Wavelet Feature Extraction and Bio-Inspired Feature Selection for the Prognosis of Lung Cancer: A Statistical Framework Analysis”. In this system Combines wavelet-based dimensionality reduction and bio-inspired feature selection to classify lung cancer subtypes Adenocarcinoma and Mesothelioma, Dragonfly algorithms is used as classifier and achieved 97.23% accuracy [12].

Proposed Methodology:

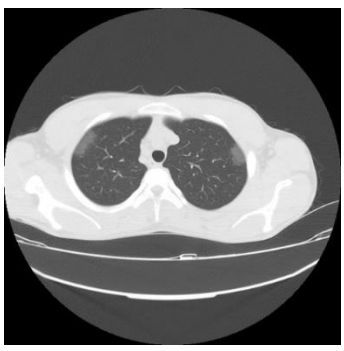


Database: The real time affected CT-Scan images of Lung Cancer, Pneumonia and COVID-19 are collected from MGM Hospital, Aurangabad, Maharashtra. Total 266 affected images were collected and categorized into mild (43 images), moderate (82 images), and severe (141 images) cases.

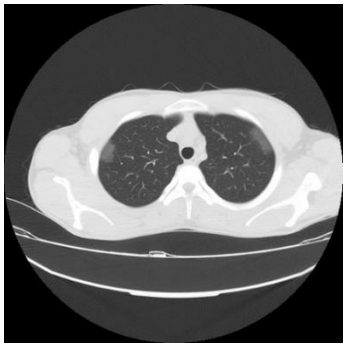
Preprocessing: Preprocessing is process that manipulate raw image into meaningful image by enhancing quality of image, resize image, noise reduction and normalization to make image ready for deep learning for training and classification model. Resizing image is necessary for standardization of dataset to uniform dimensions (520x520). Normalization used to adjust intensity values to ensure uniform contrast and brightness of input images.

Noise Reduction: Removing unwanted artifacts using smoothing, blurring and filtering

technique The method of pre-processing helps to enhance the visual appearance and quality of an image



CT Scan of Normal Lung



CT Scan Of Affected Lung

Synthetic Minority Oversampling Technique (SMOTE): Dataset collected from hospital is imbalanced, To Improve the performance of DNN model and make our three classes balanced SMOTE is a technique is used. SMOTE is a Preprocessing technique where the synthetic samples are generated for the minority class. It Reduces overfitting caused by simply duplicating minority class samples. Improves classifier performance, especially recall and F1-score.

Feature Extraction: Feature extraction is a vital step in image processing and machine learning that includes transforming locally collected data into significant illustrations that enhance image which then used in recognising different pattern and classification. Feature extraction and segmentation are initial techniques of an image processing which helps to identify complex structures, textures, edges and removes unwanted data from image. Common techniques include Edge detection (Canny and Sobel), texture feature (Histogram of oriented Gradients, Local Binary Pattern, GLCM, wavelets, Run-Length matrix), shape features (Area, Perimeter, Eccentricity, Solidity and compactness), statistical methods (Mean Intensity, Variance, Skewness, Kurtosis, Range, Entropy and Intensity Histogram). Feature extraction improves segmentation accurateness and performance of model. In this study total 30 features are extracted from affected lung image to analyse texture (18), shape (4) and edges (2) these features have been divided into two parts Wavelet packet decomposition and complementary features.

WPD is highly suitable for lung disease classification for the following reasons:

• **Texture and Pattern Analysis:** Lung diseases, such as fibrosis, emphysema, and cancer, often manifest as changes in lung texture. WPD helps in detecting these texture changes by analyzing the image at multiple resolutions.

• **Frequency Domain Representation:** Disease-related features such as nodules or lesions can be better understood in the frequency domain. WPD captures both high-frequency details (e.g., sharp edges of nodules) and low-frequency patterns (e.g., spread of lesions).

• **Directional Features:** Lung diseases might show directional characteristics due to tissue structures or disease spread patterns. WPD's horizontal, vertical, and diagonal decompositions allow for the extraction of features that highlight these directional changes.

Wavelet Packet Decomposition (WPD) Features

1. WPD hdh (Horizontal → Diagonal → Horizontal decomposition)

2. WPD vdd (Vertical → Diagonal → Diagonal decomposition)

3. WPD haa (Horizontal → Horizontal → Horizontal decomposition)

4.WPD avh (Approximation → Vertical → Horizontal decomposition)

5. WPD hah (Horizontal → Approximation → Horizontal decomposition)

6. WPD vav (Vertical → Approximation → Vertical decomposition)

7. WPD vhv (Vertical → Horizontal → Vertical decomposition)

8. WPD hhd (Horizontal → Horizontal → Diagonal decomposition)

9. WPD vdv (Vertical → Diagonal → Vertical decomposition)

10. WPD adv (Approximation → Diagonal → Vertical decomposition)

11. WPD hvd (Horizontal → Vertical → Diagonal decomposition)

12. WPD hdd (Horizontal → Diagonal → Diagonal decomposition)

13. WPD vhd (Vertical → Horizontal → Diagonal decomposition)

14. WPD aha (Approximation → Horizontal → Approximation decomposition)

15. WPD vva (Vertical → Vertical → Approximation decomposition)

16. WPD ahd (Approximation → Horizontal → Diagonal decomposition)

17. WPD vda (Vertical → Diagonal → Approximation decomposition)

18. WPD had (Horizontal → Approximation → Diagonal decomposition)

The steps to extract WPD features from lung disease images are:

1. **Preprocessing**: Convert the lung image (CT or X-ray) to grayscale if necessary, and apply denoising techniques to improve feature extraction.

2. **Wavelet Packet Decomposition**: Apply WPD to the image using a specific wavelet function (e.g., Daubechies, Symlets, Coiflets). The image is decomposed into approximation and detail coefficients, and this process is repeated for multiple levels of decomposition.

3. **Feature Calculation**: For each subband (e.g., WPD hdh, WPD vdd), calculate statistical measures such as mean, variance, entropy, and energy. These statistical measures serve as features representing the texture and frequency characteristics of the lung tissue.

Wavelet Packet Decomposition (WPD) Features in Lung Disease Classification

Wavelet Packet Decomposition (WPD) is an advanced signal processing technique used to break down an image or signal into multiple subbands of different frequency components. Unlike traditional wavelet transforms, which only decompose the approximation (low-frequency) coefficients, WPD decomposes both approximation and detail coefficients, providing a more comprehensive frequency representation. Each feature in WPD corresponds to specific subbands, capturing intricate spatial and frequency characteristics.

For instance:

• WPD hdh means horizontal filtering followed by diagonal and horizontal decomposition.

• WPD vdd refers to vertical followed by two diagonal decompositions.

By capturing fine variations in image texture and frequency, WPD enables detailed analysis of medical images such as lung disease images, helping to identify subtle abnormalities.

WPD is crucial for medical image analysis, including lung disease classification, due to its ability to:

• **Multiresolution Analysis**: Analyze the image at different resolutions and frequencies, detecting both large-scale patterns (e.g., shapes, lesions) and fine details (e.g., textures, edges).

• **Directional Sensitivity**: Decomposing the image into horizontal, vertical, and diagonal components helps to capture directionally significant features, which can correlate with disease-specific patterns in lung tissues.

• **Noise Reduction**: By focusing on specific frequency subbands, WPD can help to suppress noise while enhancing important diagnostic features.

In lung disease images (e.g., CT or X-rays), WPD is effective for identifying patterns such as irregularities in the lung texture, lesion formations, or fibrosis, which manifest differently across spatial frequencies.

Wavelet Packet Decomposition uses a combination of low-pass (h) and high-pass (g) filters applied to both approximation and detail coefficients. The equations for the decomposition at each level are:

$$A[n] = \sum_{k} x[k]h[2n-k] \quad \text{(Low-pass filter)} \qquad \ldots (1)$$

$$D[n] = \sum_{k} x[k]g[2n-k] \quad \text{(High-pass filter)} \qquad \ldots (2)$$

Where:

• A[n] is the approximation coefficients (low-frequency subband).

• D[n] is the detail coefficients (high-frequency subband).

• h and g are the wavelet filters, and x[k] is the signal or image pixel values.

At each level of decomposition, both approximation and detail coefficients are recursively decomposed to capture all frequency components.

For each subband decomposition, you calculate various statistical features. Some of the common equations include:

• Mean of the coefficients for subband S:

$$\mu_S = \frac{1}{N}\sum_{i=1}^{N} S[i]$$

• Variance of the subband coefficients:

$$\sigma_S^2 = \frac{1}{N}\sum_{i=1}^{N} (S[i] - \mu_S)^2$$

• Energy of the subband coefficients:

$$E_S = \sum_{i=1}^{N} S[i]^2$$

• Entropy of the subband coefficients:

$$H_S = -\sum_{i=1}^{N} P(S[i]) \log\big(P(S[i])\big)$$

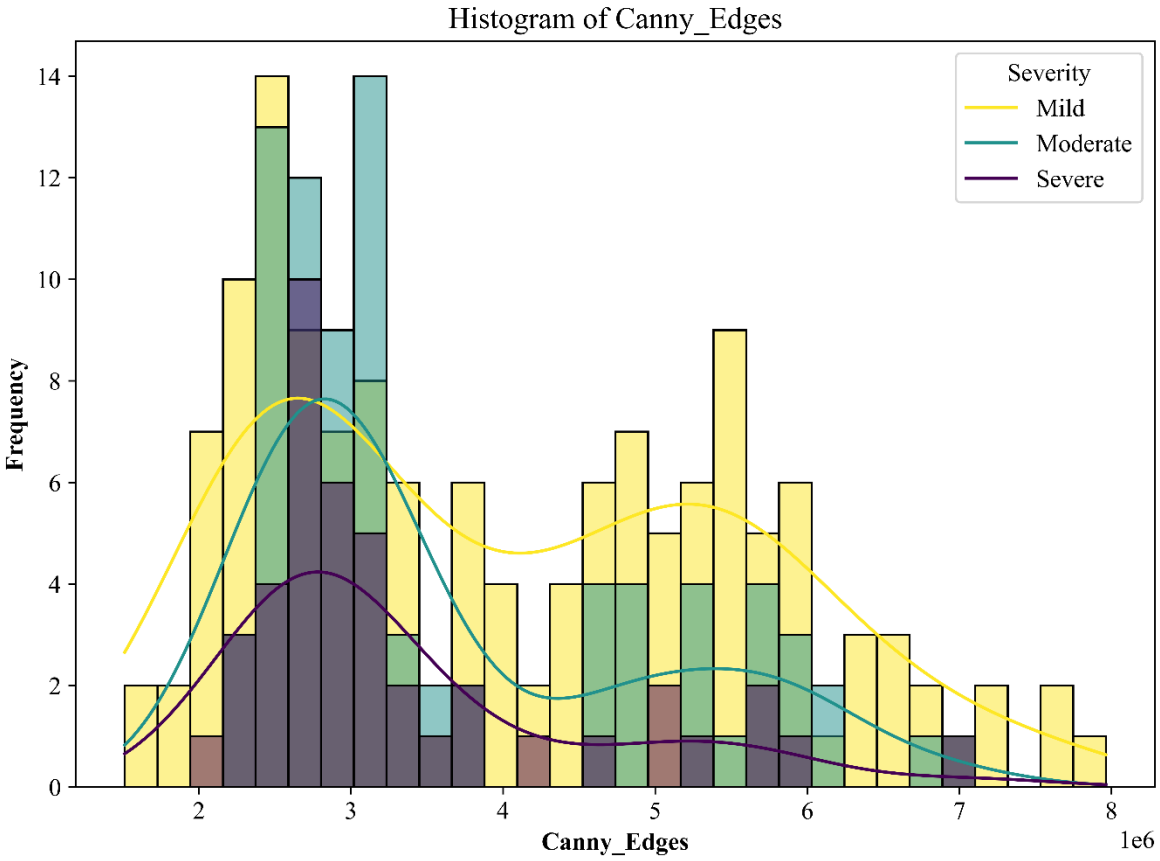
Where $P(S[i])$ is the probability of each coefficient value in the subband.

These features are then combined into a feature vector representing the lung image, which can be used in a classification model to differentiate between healthy and diseased lungs.

Edge Detection:

Edge detection features are vital in medical image analysis, especially for lung disease assessment, as they highlight the boundaries and transitions between different tissue types. Two significant edge detection features are Canny Edges and Sobel Edges X.

Canny Edges Definition: Canny Edge Detection is a multi-stage algorithm that identifies the edges in an image by detecting the points at which the image brightness changes sharply. It is renowned for its effectiveness in preserving important structural properties while reducing noise.

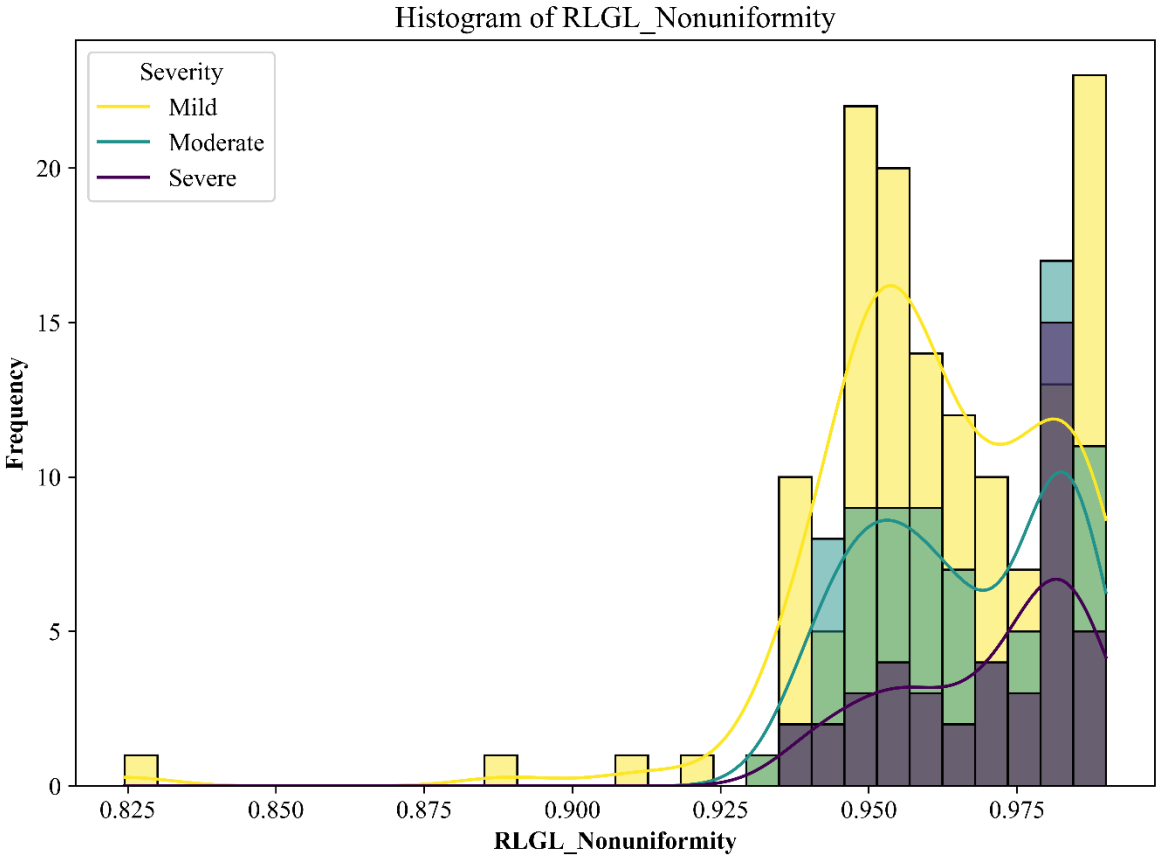


Sobel Edges X: Sobel Edge Detection is a technique that uses convolution with Sobel kernels to detect edges in an image.

Texture Features:

Run-Length Grey Level Nonuniformity (RLGN): Run-Length Grey Level Nonuniformity is a texture feature derived from the Gray Level Run Length Matrix (GLRLM). It measures the uniformity of the grey level values of pixels in runs, indicating how pixel intensities are distributed across various lengths.

Neighbor Texture Mean (NTM): Neighbor Texture Mean is a feature that quantifies the average intensity value of neighboring pixels in an image, providing insights into the local texture of lung tissues.



Shape Features:

In Lung Disease Image Analysis Shape features play a crucial role in the analysis of lung disease images, as they help in characterizing and distinguishing between different lung conditions. Here are four significant shape features: Compactness, Centroid y, Perimeter, and Number of Connected Components.

Area: The area of the largest contour detected in the binary image, indicating the size of the object.

Perimeter: The total length of the boundary of the largest contour, useful for assessing the shape's complexity.

Eccentricity: A measure of how much the shape deviates from being circular. It ranges from 0 (circular) to 1 (linear).

Solidity: The ratio of the area of the contour to the area of its convex hull. It measures how much the shape is filled.

Compactness: Calculated as Perimeter2 Area, this ratio provides a sense of how compact or spread out the shape is.

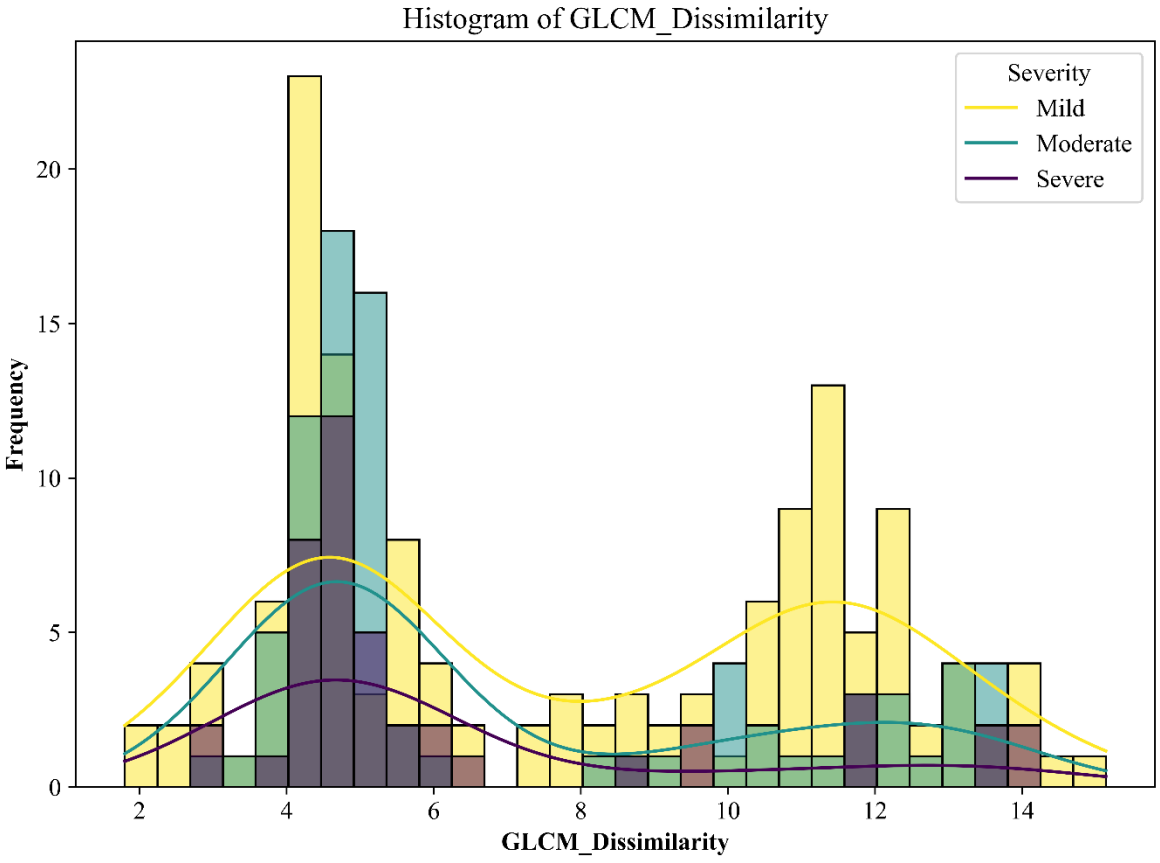
These features can be useful for various applications, such as object detection and classification.

GLCM Features:

The Gray Level Co-occurrence Matrix (GLCM) is a statistical method used to analyze the spatial relationships of pixels in an image. It captures the frequency of pixel pairs with specific values and a defined spatial relationship, which is crucial in identifying texture features. Two significant GLCM features are GLCM Correlation and GLCM Dissimilarity.

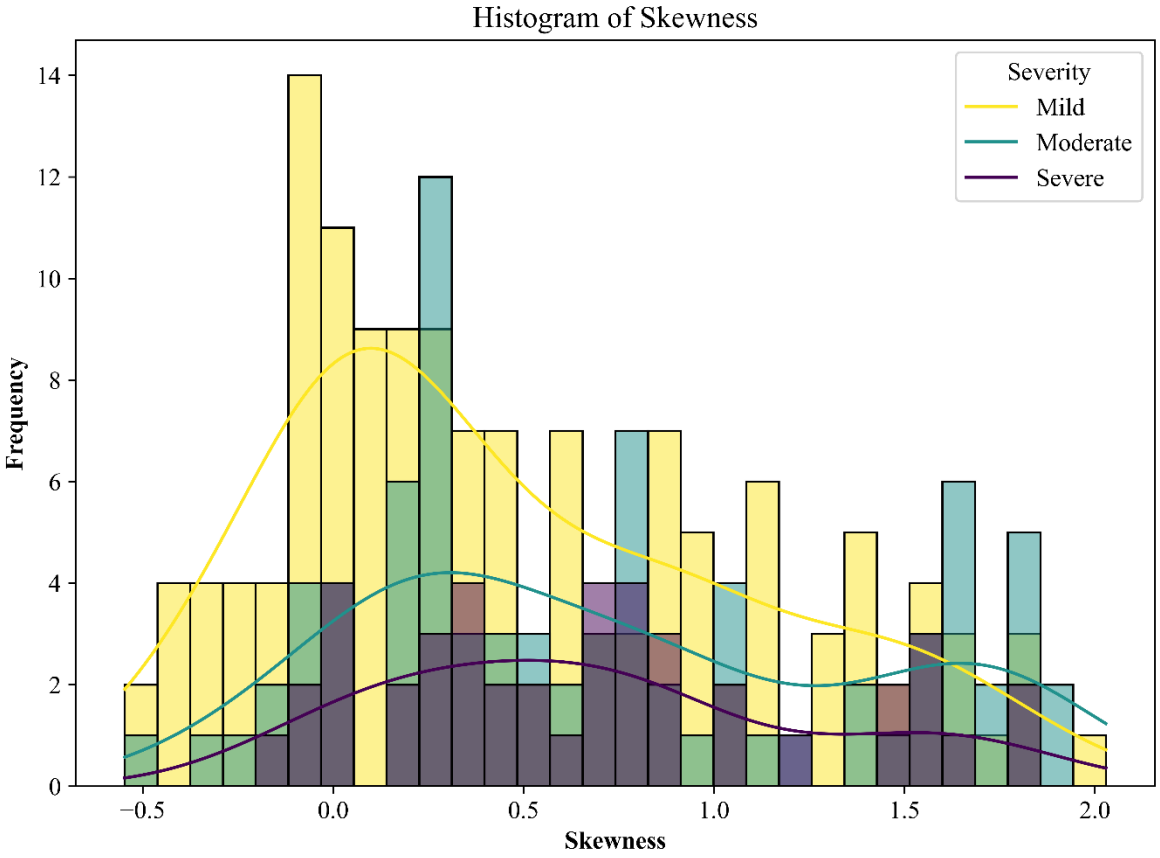
GLCM Correlation: GLCM Correlation quantifies the degree of correlation between the pixel values in the image. It measures how pixel values in a specific spatial relationship are related to each other, providing insight into the texture and structural patterns of lung tissue.

GLCM Dissimilarity: GLCM Dissimilarity measures the contrast between pixel values in the GLCM, indicating how different or dissimilar the values are within a given neighborhood. This feature is crucial for assessing the heterogeneity of the lung tissue.

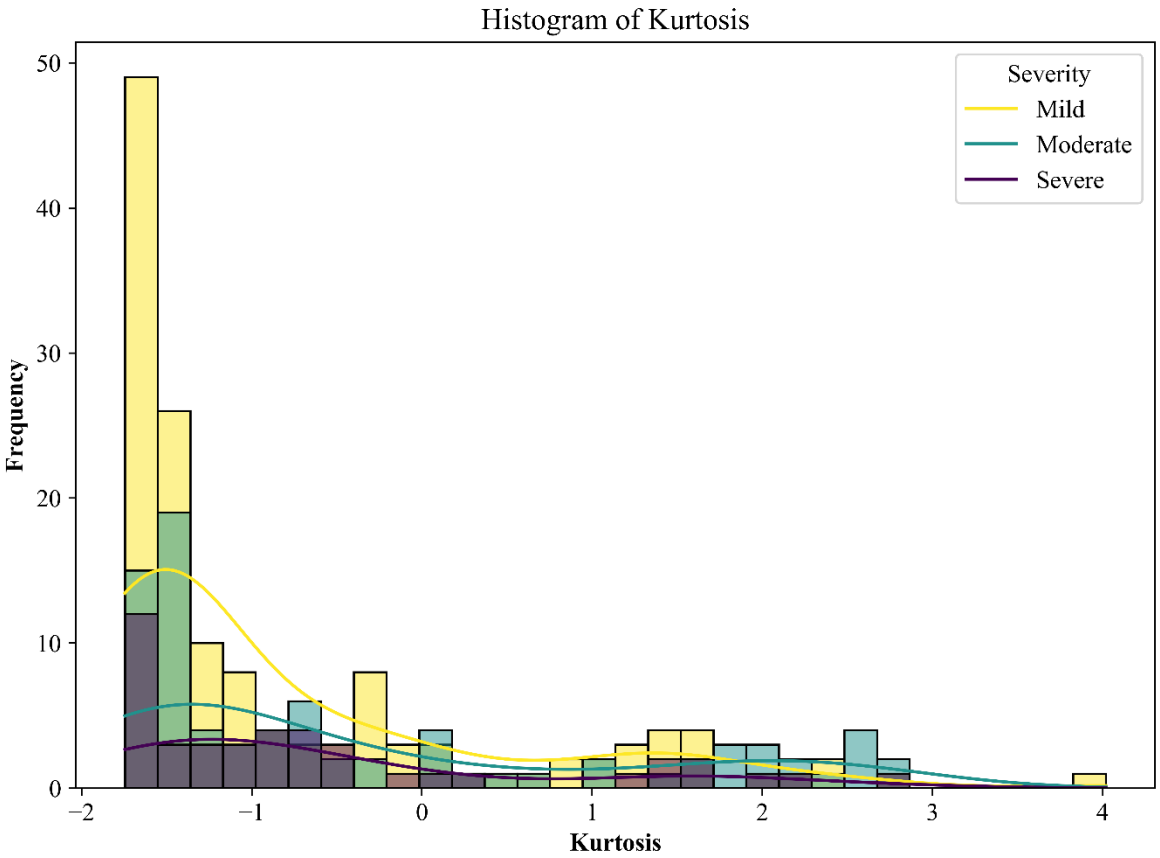


Statistical Methods:

Skewness: Skewness measures the asymmetry of the distribution of values in an image or dataset around its mean. If the distribution is symmetrical, the skewness is zero. A positive skewness indicates a distribution with a longer or fatter right tail, while a negative skewness indicates a longer or fatter left tail.



Kurtosis: Kurtosis measures the ”tailedness” or peakedness of the distribution. It in dicates whether the data are heavy-tailed or light-tailed relative to a normal distribution. A higher kurtosis suggests that data have heavy tails (outliers), while lower kurtosis implies light tails (less extreme values). In medical imaging, kurtosis helps detect abnormal distributions in lung tissues.



Training Model

Introduction to Deep Neural Networks: A Deep Neural Network (DNN) is a type of artificial intelligence that helps computers learn from data. Similar to how our brains process information through interconnected neurons, a DNN has multiple layers that transform input data into meaningful predictions. In this context, we are using a DNN to classify lung disease data into three categories: Mild, Moderate, and Severe. By analyzing various features of the data, the DNN learns to predict the appropriate disease severity based on pat terns in lung images.

Structure of the DNN: Layers in the Model

1. Input Layer: The first layer of the network takes the features (input data) from the lung disease dataset. Each feature corresponds to an aspect of the lung images (e.g., texture, edge detection, etc.).

2. Hidden Layers: The model consists of four hidden layers, where the actual learning occurs
• First Hidden Layer: Contains 256 neurons. Each neuron receives input from the previous layer and uses a function called ReLU (Rec tified Linear Unit) to determine its output. ReLU helps introduce non-linearity, allowing the network to learn complex relationships in the data.

• Second Hidden Layer: Contains128 neurons and also uses ReLU activation.

• Third Hidden Layer: Contains 64 neurons, applying the same activation function.

• Fourth Hidden Layer: Contains 32 neurons.

3. Output Layer: The final layer consists of 3 neurons, each representing one of the lung disease severity categories (Mild, Moderate, Severe). It uses a softmax activation function, which converts the outputs into prob abilities that sum to one. This allows the model to predict the likelihood of each class for a given input. To improve the learning we introduced technique Batch Normalization (Applied after each hidden layer, to help model learn more effectively) and Dropout (added after each batch normalization layer to prevents the model from becoming too reliant on any single neuron, helping to avoid overfitting)

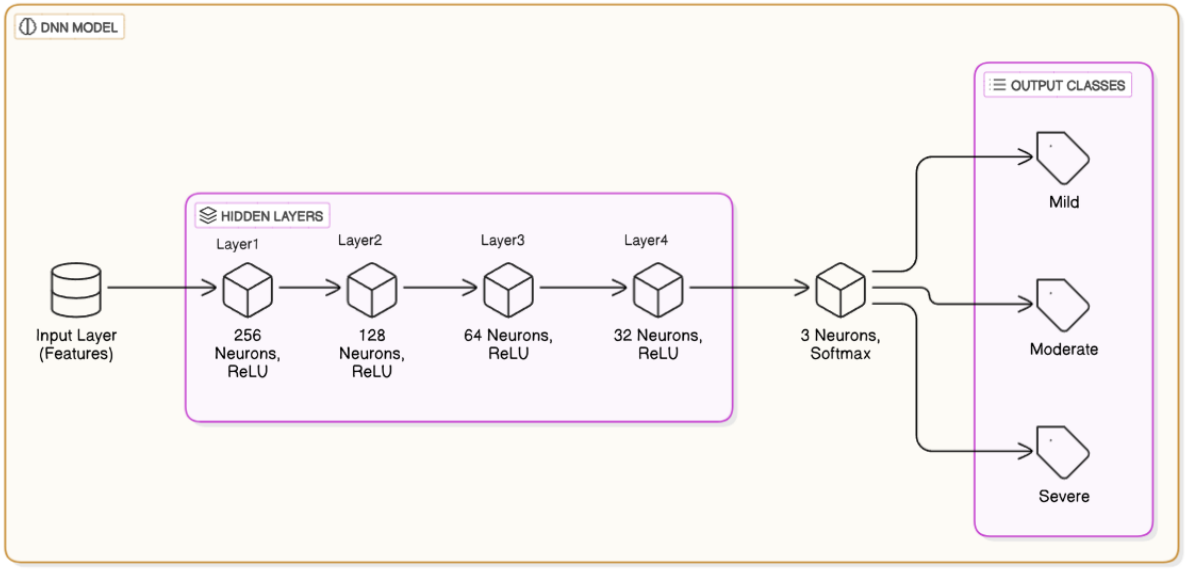


Fig: Diagrammatic Structure of DNN Model

Training Process The model is set to train for a maximum of 500 epochs, meaning it will go through the entire dataset 500 times. A batch size of 32 is used, indicating that the model will process 32 samples at a time before updating the weights.

Callback: To enhance the training process, two important callbacks are used:

1. Early Stopping: This callback monitors the validation loss and stops training if it does not improve for a certain number of epochs (20 epochs). This helps prevent overfitting by ensuring the model does not continue to train once it has learned as much as possible from the data.

2. Reduce Learning Rate on Plateau: This callback reduces the learning rate by half if the validation loss does not improve for 10 epochs. This allows the model to make finer adjustments as it approaches a minimum, improving overall performance.

Classification

A classification report provides key metrics to evaluate the performance of a model

• **Precision:** The proportion of true positive predictions to the total predicted positives. It indicates how many selected items are relevant.

• **Recall:** The proportion of true positives to the total actual positives. It indicates how many relevant items are selected.

• **F1-score:** The harmonic mean of precision and recall, giving a balance between the two.

Training Classification Report:

• Mild: Precision = 0.95, Recall = 0.96, F1-score = 0.95

• Moderate: Precision = 0.95, Recall = 0.92, F1-score = 0.93

• Severe: Precision = 0.94, Recall = 0.96, F1-score = 0.95

Overall accuracy for training is approximately 95%.

Validation Classification Report:

•Mild: Precision = 0.81, Recall = 1.00, F1-score = 0.89

•Moderate: Precision = 0.94, Recall = 0.76, F1-score = 0.84

•Severe: Precision = 1.00, Recall = 0.95, F1-score = 0.98

Overall accuracy for validation is approximately 90%.

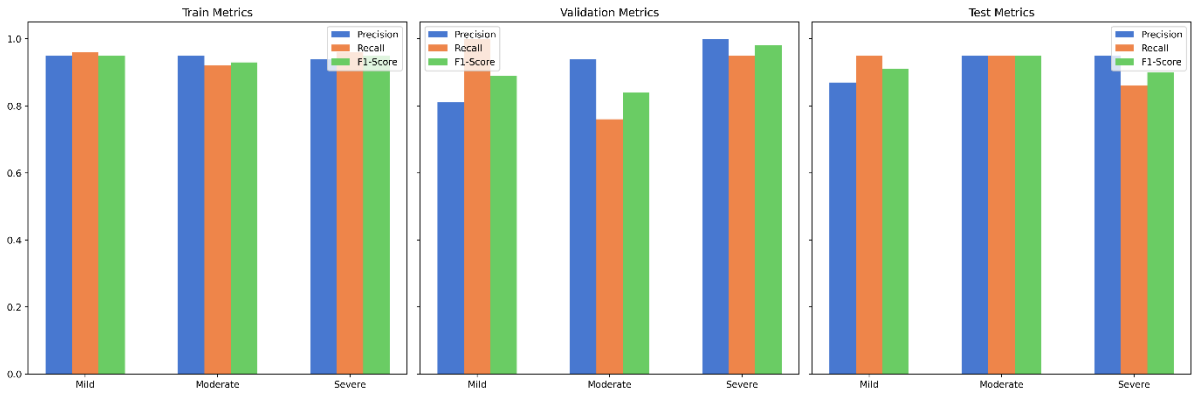
Testing Classification Report:

•Mild: Precision = 0.87, Recall = 0.95, F1-score = 0.91

• Moderate: Precision = 0.95, Recall = 0.95, F1-score = 0.95

• Severe: Precision = 0.95, Recall = 0.86, F1-score = 0.90

Overall accuracy for testing is approximately 92%.



Result Analysis:

• Confusion Matrix: This matrix shows the number of true positive, true negative, false positive, and false negative predictions for both the training and testing datasets. It helps in understanding how well the model is classifying each disease severity category.

• Classification Report: This report provides detailed metrics such as precision, recall, F1-score, and support for each class. These metrics help in assessing the model's performance across all categories.

Confusion Matrix:

A confusion matrix is a table that allows us to visualize the performance of a classification model. Each row of the matrix represents the instances in an actual class, while each column represents the instances in a predicted class

Train Set

$$\begin{bmatrix} 95 & 3 & 1 \\ 3 & 91 & 5 \\ 2 & 2 & 94 \end{bmatrix}$$

Training classification report:

Table :Classification Report for Training Set

Class	Precision	Recall	F1-Score	Support
Mild	0.95	0.96	0.95	99
Moderate	0.95	0.92	0.93	99
Severe	0.94	0.96	0.95	98
Accuracy		0.95		296
Macro Avg	0.95	0.95	0.95	296
Weighted Avg	0.95	0.95	0.95	296

Validation Set

$$\begin{bmatrix} 21 & 0 & 0 \\ 5 & 16 & 0 \\ 0 & 1 & 20 \end{bmatrix}$$

Validation Classification Report:

Table: Classification Report for Validation Set

Class	Precision	Recall	F1-Score	Support
Mild	0.81	1.00	0.89	21
Moderate	0.94	0.76	0.84	21
Severe	1.00	0.95	0.98	21
Accuracy		0.90		63
Macro Avg	0.92	0.90	0.90	63
Weighted Avg	0.92	0.90	0.90	63

Test set

$$\begin{bmatrix} 20 & 0 & 1 \\ 1 & 20 & 0 \\ 2 & 1 & 19 \end{bmatrix}$$

Test Classification

Table: Classification Report for Test Set

Class	Precision	Recall	F1-Score	Support
Mild	0.87	0.95	0.91	21
Moderate	0.95	0.95	0.95	21
Severe	0.95	0.86	0.90	22
Accuracy		0.92		64
Macro Avg	0.92	0.92	0.92	64
Weighted Avg	0.92	0.92	0.92	64

Overall Accuracy:

•Train Accuracy: 0.9459

•Validation Accuracy: 0.9048

•Test Accuracy: 0.9219

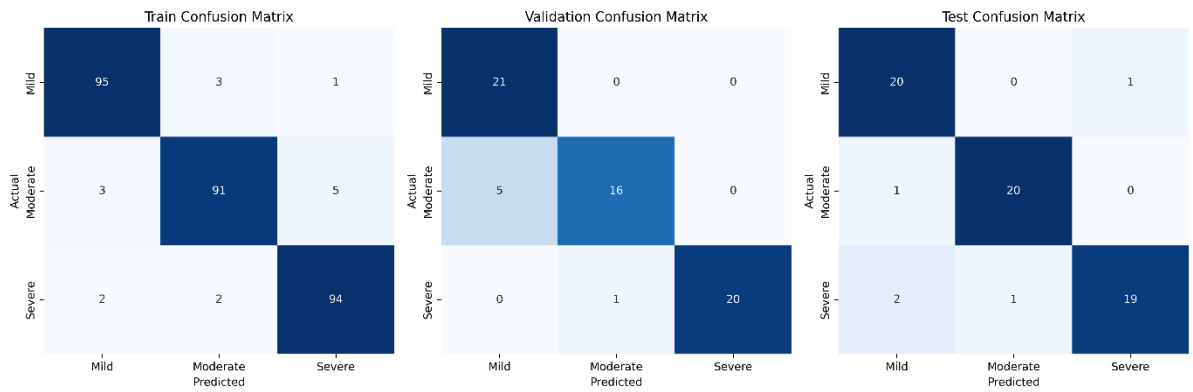
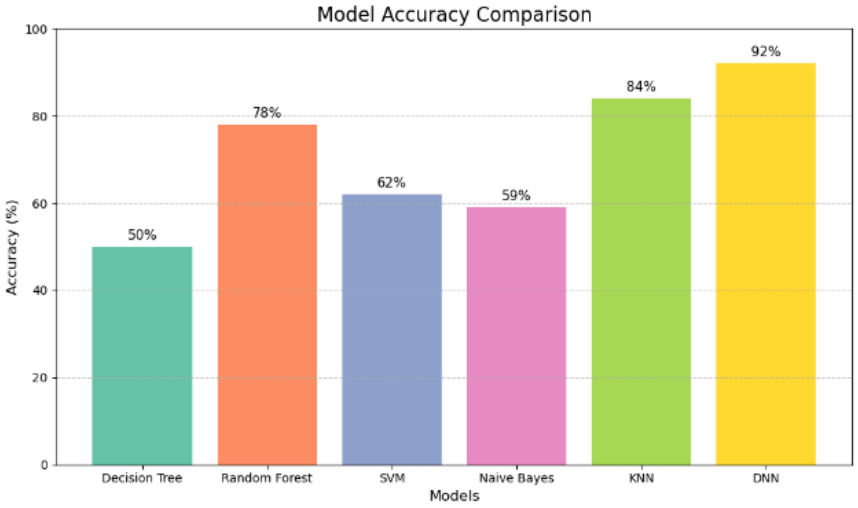


Fig: Confusion Matrix Training, Validation and Testing

Comparative Analysis:

After completing the preprocessing of medical image data collected from the hospital, the processed dataset was used to train a Deep Neural Network (DNN) model. The performance of this model was then compared with other models to evaluate its effectiveness. The evaluation was conducted using key performance metrics such as accuracy, precision, recall, F1-score. This comparison helped in determining whether the proposed DNN model outperformed traditional machine learning methods or other deep learning architectures Decision Tree, Random Forest, SVM, Naïve Bayes and KNN in accurately classifying or segmenting the medical images.



In addition, to evaluate the generalization performance of the proposed model beyond hospital-acquired data, we also tested it using publicly available online medical image datasets, COVID-19-CT-Dataset (Kaggle). The primary objective of this comparison was to analyze how accurately the model can classify images into the three predefined diagnostic categories: Mild, Moderate and Severe. The model successfully categories online dataset into the three categories with its severity.

Conclusion:

This Study Focuses on Detection and classification of respiratory diseases in accordance to the severity (Mild, Moderate and Severe). CT scan Images consist of diseases like Lung Cancer, Covid-19 and Pneumonia. To address the class imbalance in database, the synthetic minority oversampling technique (SMOTE) was applied. Total of 30 multi-domain features ware extracted using a comprehensive feature extraction framework by integrating wavelet packet decomposition (WPD) with statistical texture, shape, edge detection and Gray level co-occurrence matrix (GLCM) features. Further Deep Neural Network Model is trained for the classification on the basis of severity. 70% of data is used for training, 15% for Validation and 15% for Testing purpose, and achieved 92% accuracy.

Chapters Scheme:

Chapter 1: Introduction

Chapter 2: Literature review

Chapter 3: Dataset Collection

Chapter 4: Proposed Methodology and Implementation

Chapter 5: Result Analysis

Chapter 6: Conclusion and Future work.

Name and Signature of Research Guide **Name and Signature of Candidate**

(Dr. Prapti D. Deshmukh) **(Urvashi B. Deshmukh)**

Principal,

Dr. G.Y. Pathrikar College of CS and IT,

Aurangabad, MS (India)

References:

1. Smith, A., Johnson, M., & Lee, J. (2021). Wavelet-based feature extraction for medical image analysis. IEEE Transactions on Biomedical Engineering, 68(2), 145–157.
2. Jones, P., & Patel, R. (2020). Texture analysis for lung disease classification using GLCM. Journal of Medical Imaging, 12(4), 250–260.
3. Zhao, Y., Chang, T., & Li, X. (2019). Role of statistical features in medical image segmentation. International Journal of Computerized Medical Imaging and Graphics, 33(1), 45–60.
4. Liu, J., & Wang, K. (2020). Shape features for robust lung disease classification. Computerized Medical Imaging and Graphics, 44(3), 150–165.
5. Kieu, Stefanus T.H., Abdullah Bade, Mohd H.A. Hijazi, and Hoshang Kolivand. 2020. "A Survey of Deep Learning for Lung Disease Detection on Medical Images: State-of-the-Art, Taxonomy, Issues and Future Directions" Journal of Imaging 6, no. 12: 131. https://doi.org/10.3390/jimaging6120131
6. A. D. Gunasinghe, A. C. Aponso and H. Thirimanna, "Early Prediction of Lung Diseases," 2019 IEEE 5th International Conference for Convergence in Technology (I2CT), 2019, pp. 1-4, doi: 10.1109/I2CT45611.2019.9033668.
7. Tianyi Li, Wei Wei, Lidan Cheng, Shengjie Zhao, Chuanjun Xu, Xia Zhang, Yi Zeng, Jihua Gu, "Computer-Aided Diagnosis of COVID-19 CT Scans Based on Spatiotemporal Information Fusion", Journal of Healthcare Engineering, vol. 2021, Article ID 6649591, 11 pages, 2021. https://doi.org/10.1155/2021/6649591
8. Yuanyuan Peng, Hualan Zhong, Zheng Xu, Hongbin Tu, Xiong Li, Lan Peng, "Pulmonary Lobe Segmentation in CT Images Based on Lung Anatomy Knowledge", Mathematical Problems in Engineering, vol. 2021, Article ID 5588629, 15 pages, 2021. https://doi.org/10.1155/2021/5588629
9. Suren Makaju, P.W.C. Prasad, Abeer Alsadoon, A.K. Singh, A. Elchouemi, “Lung Cancer Detection using CT Scan Images” Procedia Computer Science, Volume 125, 2018, Pages 107-114, ISSN 1877-0509, https://doi.org/10.1016/j.procs.2017.12.016
10. X. Jin, Y. Zhang and Q. Jin, "Pulmonary Nodule Detection Based on CT Images Using Convolution Neural Network," 2016 9th International Symposium on Computational Intelligence and Design (ISCID), 2016, pp. 202-204, doi: 10.1109/ISCID.2016.1053.
11. S. Bakheet and A. Al-Hamadi, ‘Automatic detection of COVID-19 using pruned GLCM-Based texture features and LDCRF classification’, *Computers in Biology and Medicine*, vol. 137, p. 104781, 2021. https://doi.org/10.1016/j.compbiomed.2021.104781
12. Karthika, M. S., Rajaguru, H., & Nair, A. R. (2024). Wavelet feature extraction and bio-inspired feature selection for the prognosis of lung cancer—A statistical framework analysis. Measurement, 238, 115330. https://doi.org/10.2139/ssrn.4697036
13. Synthesizing Radiological Insights: Enhancing Lung Disease Classification through Multimodal Imaging. Kumari CS, Seethalakshmi K.

List of Published Papers

1. U. B. Deshmukh, A. S. Solanke, and P. D. Deshmukh, 'Recent Study of Lung Disease Detection Using Deep Learning Techniques', in Proceedings of the International Conference on Applications of Machine Intelligence and Data Analytics (ICAMIDA 2022), 2023, pp. 755–760.
2. U. B. Deshmukh and P. D. Deshmukh, 'Comparative Analysis Of COVID-19 Images Through Different Filtering Techniques Using PSNR, MSE And SSIM', in Hinweis Third International Conference on Artificial Intelligence and Data Science (AIDE)-24. indexed by Scopus as conference papers
3. U. B. Deshmukh and P. D. Deshmukh, 'Comparative Analysis Of COVID-19 Images Through Different Filtering Techniques Using PSNR, MSE And SSIM' in the Grenze International Journal of Engineering and Technology (GIJET) as Volume 11 Issue 1 (January 2025 issue)
4. U. B. Deshmukh and P. D. Deshmukh, 'FRAMEWORK FOR COMPREHENSIVE FEATURE EXTRACTION FORMEDICAL IMAGE ANALYSIS USING WAVELET PACKETDECOMPOSITION AND COMPLEMENTARY DESCRIPTORS', in World Journal of Engineering Research and Technology, VOLUME 11, FEBRUARY ISSUE 2

Communicated and Accepted Papers

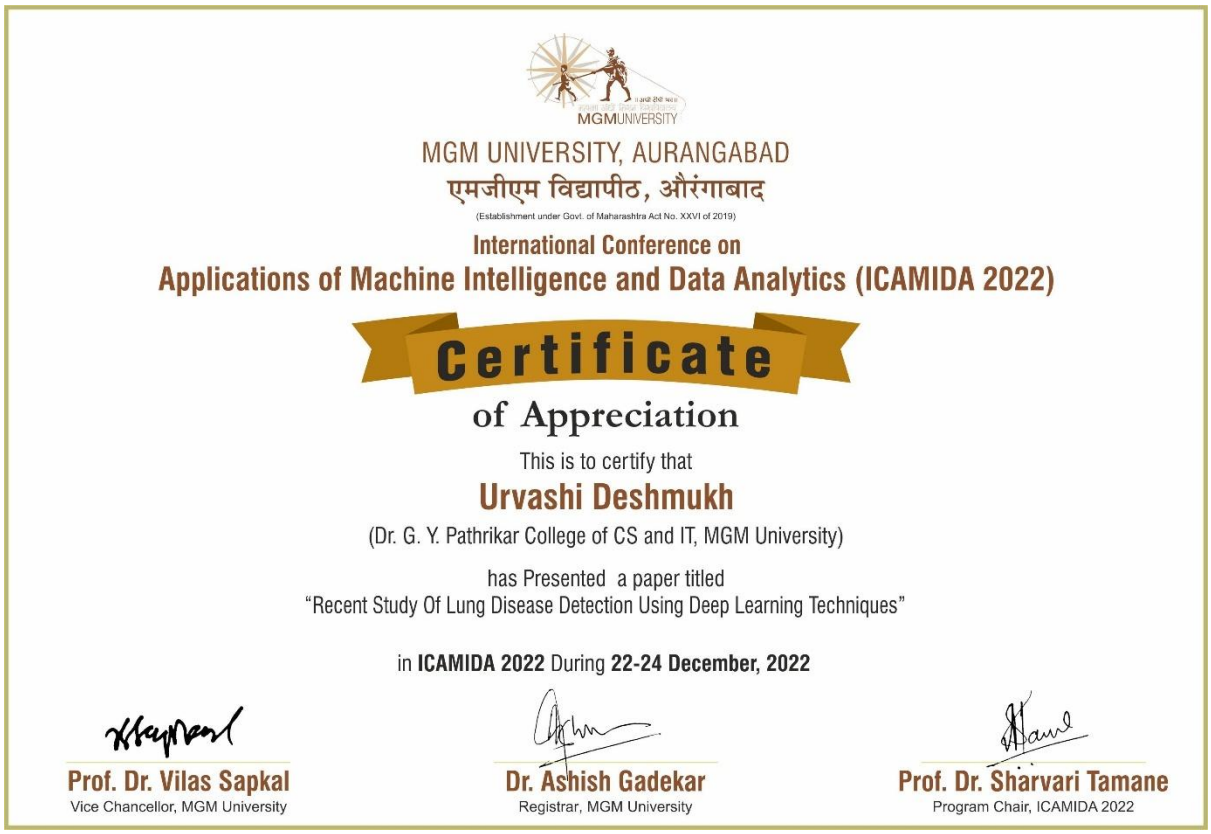
1. Accepted 'Advancing Diagnostic Accuracy in Lung Disease Severity Classification Using Multi-Domain Features' , in International Journal of Preventive Medicine and Health

Papers as Co-Author:

1. Apurva Solanke, Urvashi Deshmukh, & Prapti Deshmukh (2023). Review of the Study of Automated Skin Cancer Detection Using Digital Image Processing. In Proceedings of the International Conference on Applications of Machine Intelligence and Data Analytics (ICAMIDA 2022) (pp. 980-988). Atlantis Press.

Proof of Publication:

U. B. Deshmukh, A. S. Solanke, and P. D. Deshmukh, 'Recent Study of Lung Disease Detection Using Deep Learning Techniques', in *Proceedings of the International Conference on Applications of Machine Intelligence and Data Analytics (ICAMIDA 2022)*, 2023, pp. 755–760.



MGM UNIVERSITY, AURANGABAD

एमजीएम विद्यापीठ, औरंगाबाद

(Establishment under Govt. of Maharashtra Act No. XXVI of 2019)

International Conference on

Applications of Machine Intelligence and Data Analytics (ICAMIDA 2022)

Certificate

of Appreciation

This is to certify that

Urvashi Deshmukh

(Dr. G. Y. Pathrikar College of CS and IT, MGM University)

has Presented a paper titled

"Recent Study Of Lung Disease Detection Using Deep Learning Techniques"

in **ICAMIDA 2022** During **22-24 December, 2022**

Prof. Dr. Vilas Sapkal
Vice Chancellor, MGM University

Dr. Ashish Gadekar
Registrar, MGM University

Prof. Dr. Sharvari Tamane
Program Chair, ICAMIDA 2022

U. B. Deshmukh and P. D. Deshmukh, ‘Comparative Analysis Of COVID-19 Images Through Different Filtering Techniques Using PSNR, MSE And SSIM’, in *Hinweis Third International Conference on Artificial Intelligence and Data Science (AIDE)-24.* indexed by Scopus as conference papers

CERTIFICATE OF PRESENTATION & PUBLICATION

THE HINWEIS RESEARCH

Hinweis Third International Conference on Artificial Intelligence and Data Science, AIDE-2024

November 30, 2024

http://icaids.thehinweis.com/2024

Urvashi B Deshmukh

Dr. GY Pathrikar College of Computer Science and IT, Aurangabad, Maharashtra, India

Author of a Research Paper titled ***Comparative Analysis of COVID-19 Images through Different Filtering Techniques using PSNR, MSE and SSIM*** has submitted the paper which has been approved and presented for publication in the Hinweis Third International Conference on Artificial Intelligence and Data Science, AIDE-2024.

Dr. Yogesh Chaba
General Chair

Hinweis Research

Dr. Janahanlal Stephen
General Co Chair

U. B. Deshmukh and P. D. Deshmukh, ‘Comparative Analysis Of COVID-19 Images Through DifferentFiltering Techniques Using PSNR, MSE And SSIM’ in the Grenze International Journal of Engineering and Technology (GIJET) as Volume 11 Issue 1 (January 2025 issue)

Grenze International Journal of Engineering and Technology, January Issue

Comparative Analysis of COVID-19 Images through different Filtering Techniques using PSNR, MSE and SSIM

Ms. Urvashi B. Deshmukh[1] and Dr. Prapti D. Deshmukh[2]
[1-2]Dr. G. Y. Pathrikar college of Computer Science and IT, MGM University, Chhatrapati Sambhaji Nagar, India
Email: urvashi.deshmukh17@gmail.com, prapti.research1@gmail.com

***Abstract*— Lung diseases are a significant global health concern, and early detection is imperative due to their contagious nature.**
Integrating computer-aided detection technology with im-aging techniques can significantly enhance diagnostic accuracy.
Preprocessing techniques are pivotal for refining image quality.
Research has been conducted to identify the most suitable preprocessing filters for COVID-19 images obtained from hospitals. Evaluation techniques such as Peak Signal-To-Noise Ratio (PSNR), Structured Similarity Index Meth-od (SSIM), and Mean Square Error (MSE) are utilized to assess image quality.
This study aims to determine the optimal techniques for enhancing image quality and improving diagnostic precision on private datasets. In conclusion, 62 images exhibited superior results with the bilateral filter, while the remaining 41 images demonstrated optimal outcomes with the Gaussian filter when analyzing images collected from the hospital.

***Index Terms*— MSE, PSNR, SSIM, Mean Filter, Gaussian Filter, Bilateral filter Laplacian Filter, Median Filter, image Preprocessing.**

I. INTRODUCTION

Lung diseases represent one of the most prevalent medical situations globally, affecting millions. These conditions encompass a range of issues that impede the lungs' ability to function correctly.
Image data preprocessing has a vital role in enhancing accuracy & reliability of subsequent analyses. Study conducts a comprehensive comparison of various filtering techniques applied to private dataset of COVID-19, which have been meticulously collected from MGM Hospital, Chhatrapati Sambhaji Na-gar.
By evaluating the effectiveness of these filtering methods, we aim to determine the most suitable techniques for optimizing image quality and facilitating more precise diagnostic outcomes, this study focus on filters for preprocessing.
After collecting the data, we meticulously reviewed it to identify images with the most significantly affected areas. where different filters has been applied such as Mean Filter, Median Filter, Gaussian Filter, Bilateral filter Laplacian Filter. From this process, we selected lung CT images that exhibited the highest degree of abnormality for analysis.
The resulting images apply the different filters and it is evaluated with PSNR, SSIM and MSE measures.

Grenze ID: 01.GIJET.11.1.10
© Grenze Scientific Society, 2025

U. B. Deshmukh and P. D. Deshmukh, ‘FRAMEWORK FOR COMPREHENSIVE FEATURE EXTRACTION FORMEDICAL IMAGE ANALYSIS USING WAVELET PACKETDECOMPOSITION AND COMPLEMENTARY DESCRIPTORS’, in World Journal of Engineering Research and Technology, VOLUME 11, FEBRUARY ISSUE 2

WORLD JOURNAL OF ENGINEERING RESEARCH AND TECHNOLOGY
www.wjert.org editor@wjert.org Impact Factor 7.029
ISSN 2454-695x
ISO 9001: 2015 Certified Journal
PUBLICATION CERTIFICATE
This is to certify that Article entitled
"FRAMEWORK FOR COMPREHENSIVE FEATURE EXTRACTION FOR MEDICAL IMAGE ANALYSIS USING WAVELET PACKET DECOMPOSITION AND COMPLEMENTARY DESCRIPTORS"
With Manuscript no. WJERT/2838/11/2025 By
Urvashi B. Deshmukh*
has been published in VOLUME 11, FEBRUARY ISSUE 2 after getting reviewed by three reviewers.
EDITOR IN CHIEF
WJERT
MANAGING EDITOR

Apurva Solanke, Urvashi Deshmukh, & Prapti Deshmukh (2023). Review of the Study of Automated Skin Cancer Detection Using Digital Image Processing. In Proceedings of the International Conference on Applications of Machine Intelligence and Data Analytics (ICAMIDA 2022) (pp. 980-988). Atlantis Press.

Review of the Study of Automated Skin Cancer Detection Using Digital Image Processing

Apurva Solanke(✉), Urvashi Deshmukh, and Prapti Deshmukh

Department of Computer Science, MGM University, Dr. G. Y. Pathrikar College of C.S. I.T., Aurangabad, Maharashtra, India
asolankeresearch@gmail.com

Abstract. The unbalanced and inadequate distribution of medical services, for example, has grown more and more obvious in today's climate, where medical care challenges have intensified. In this instance, the use of Machine Learning has been a natural progression in the development of modern medical care. As a result of the development of cosmetics use, numerous malignancies and tumors have appeared in the world in recent years. One such dangerous cancer or tumor is melanoma. The malignant illness known as melanoma occurs when the pigment-producing cells that give the skin its color turn cancerous. Signs could include a sudden, erratic growth or a change in an existing mole. Any area of the body might acquire melanoma. Hence Using methodologies involving lesion detection for accuracy, efficiency, and performance criteria, skin lesions are automatically detected. For the purpose of early skin lesion identification, the suggested approach employs feature extraction utilizing the ABCD (Asymmetry, Border, Color, Diameter) rule, GLCM (Gray Level Co-Occurrence Matrix, and HOG (Histogram Of Oriented Gradients) feature extraction. Pre-processing is used in the proposed work by researchers to enhance the skin lesion's clarity and quality while reducing artefacts such skin color and hair. The lesion part was segmented individually using Geodesic Active Contour (GAC), which was also effective for feature extraction. To extract the properties of symmetry, border, color, and diameter, the ABCD scoring method was utilized. Textural characteristics were extracted using HOG and GLCM. Different machine learning approaches, such as SVM (Support Vector Machine), KNN (K-Nearest Neighbor), and Naive Bayes classifier, are used to categories skin lesions between benign and melanoma utilizing the retrieved features.

Keywords: Diabetic Retinopathy · Eyes · Deep learning · Machine Learning

1 Introduction

The primary organ of the human body is the skin, which serves as the body's outermost layer. The skin has up to seven layers of ectodermal tissues that shield the internal organs, muscles, and bones beneath. Antioxidants help regulate body temperature, shield the physical body from bacteria and dangerous substances, and allow for the sensations of cold, fire, and contact. When a section of the skin differs significantly from other parts

© The Author(s) 2023
S. Tamane et al. (Eds.): ICAMIDA 2022, ACSR 105, pp. 980–988, 2023.
https://doi.org/10.2991/978-94-6463-136-4_86

Accepted ‘Advancing Diagnostic Accuracy in Lung Disease Severity Classification Using Multi-Domain Features’ , in International Journal of Preventive Medicine and Health

International Journal of Preventive Medicine and Health
ISSN: 2582-7588 (online) | Exploring Innovation | A Key for Dedicated Services
Published by: Lattice Science Publication
G-20, Block A, Tirupati Abhinav Commercial Campus, Tirupati Abhinav Homes, Ayodhya Bypass Road, Damkheda, Bhopal (M.P)-462037, India
Website: www.ijpmh.latticescipub.com Email: support@latticescipub.com

Acceptance Letter

Dear Author(s): Urvashi Deshmukh, Prapti Deshmukh

Submission ID:	IJRTE_6117
Paper ID:	D107805040525
Paper Title:	Advancing Diagnostic Accuracy in Lung Disease Severity Classification Using Multi-Domain Features

The above article has been evaluated and **accepted** by the editorial board for publication in the **'International Journal of Preventive Medicine and Health (IJPMH)'**, which can be published in the regular issue of **Volume-5**, **Issue-4**, **May 2025**, if the corresponding author submits the supporting documents within the stipulated time. The journal employs a First-Come-First-Serve (FCFS) priority model for allocating articles within each volume's issues. The editor retains the right to incorporate the article in subsequent issues of the respective volume if the current issue does not have adequate capacity. **To ensure adherence, it is recommended that the required documentation be furnished electronically (online) through the specified URL by 06 April 2025:**

1. **Article Processing Charge (APC): https://www.latticescipub.com/fee/**

   Authors should process the Article Process Charge (APC) first and then submit it to the journal through point no 2 (below). As an open-access journal with no subscription charges, APC is payable by the author or research funder to cover the costs associated with publication. This ensures that the article will be published immediately and permanently, free of access to everyone.

2. **Copyright and Final Article Submission: https://www.latticescipub.com/copyright/**

   a. ***Formatting of the Article: Optional***
   The author can submit the final article (camera-ready paper) in either single-column format or formatted according to the journal template available at: www.ijpmh.latticescipub.com/download
   b. ***Header, Footers, and Dates: not required***
   The journal editors will only edit the article's headers, footers, and dates.
   c. ***Author's Profile: Required***
   The author profile is mandatory and should include a photo of the author and a minimum of 100 words about the author.
   d. ***Reference Citation and Corrections: Required***
   Authors should ensure that the references used in the reference section have been correctly cited (used) in the article. The author may improve the article according to the reviewer's comments. Authors should carefully study the "guidelines" for the preparation and formatting of the article before submission to the journal.

Mohini Sahu
[Manager, Lattice Science Publication (LSP)]

Dr. Lucy Mohapatra
(Editor-In-Chief)
[Asst. Professor, Amity Institute of Pharmacy, Lucknow (UP), India]

Website: www.ijpmh.latticescipub.com **Indexing and Abstracting:** www.ijpmh.latticescipub.com/indexing
Email: support@latticescipub.com **For any Query:** www.ijpmh.latticescipub.com/faq
Editorial and Publishing Policies: www.ijpmh.latticescipub.com/ethics-policies **DOI:** www.doi.org/10.54105/ijpmh
Guidelines for Authors: www.ijpmh.latticescipub.com/instruction-for-authors **Open Access Publishing:** www.ijpmh.latticescipub.com/open-access-license